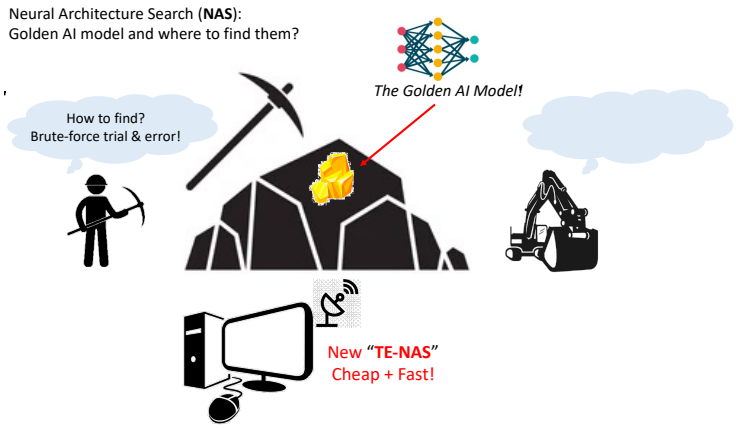
# Accelerating Neural Architecture Search with Theory-Grounded, Training-Free Metrics

Over the past decade, the world has seen tremendous increases in the deployment of artificial intelligence (AI) technology. The main horsepower behind the success of AI systems is provided by deep learning models and machine learning (ML) algorithms. Recently, a new AI paradigm has emerged: Automated Machine Learning (AutoML) including its subfield Neural Architecture Search (NAS).

State-of-the-art ML models consist of complex workflows with numerous design choices and variables that must be tuned for optimal performance. Optimizing all the variables manually can be complex and even intractable. Neural Architecture Search (NAS) allows us to find a high-performing deep learning model architecture automatically. NAS uses ML models to design or train other ML models, executing trial-and-error processes millions or billions of times faster than humans. The figure below illustrates this paradigm shift and the place of a new tool known as TE-NAS within it.

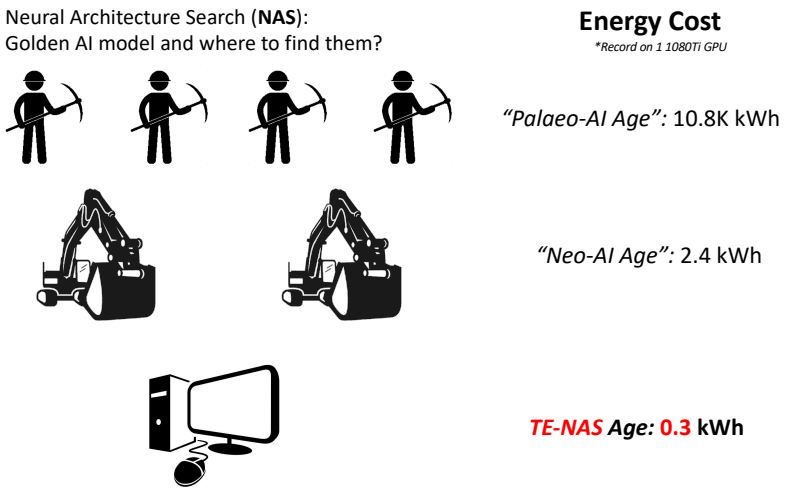



NAS automates the process of finding a good model, but this process can be quite expensive. NAS has to examine many (good and bad) models' performance before discovering the pattern for a good architecture, and it can take a long time even to determine if a single model will perform well or not. Training NAS thus requires the use of supercomputers, but even with advanced supercomputers, it can take days or even weeks.

Researchers at the University of Texas Austin have developed a new method to design NAS models and predict their success and accuracy that eliminates the high costs and delays associated with their training. The new design model called TE-NAS (and a follow-up known as As-ViT) uses two measurements grounded in recent findings in deep learning mathematical theories, which can accurately predict the performance of a deep learning model. These measurements can be used to forecast performance without expensive training runs. In just tens of seconds, TE-NAS can tell if a model is good or not. TE-NAS can thus complete NAS over hundreds of models in only 4 hours.

“Compared with previous state-of-the-art NAS methods, we reduced the search latency by up to 100 times and used only a fraction of their energy costs (illustrated in the figure below), and can still preserve the same high quality of the found models,” said Atlas Wang, Ph.D., principal investigator of the research.

“Our work has the potential to boost the discovery of AI models on many applications such as computer vision and natural language processing. With this extremely fast and lightweight NAS framework, more users can afford to customize their deep learning models, so NAS is becoming democratized. TE-NAS will allow AI practitioners to customize more affordable, more adaptable, and more usage-specific AI models.”




*An overview of TE-NAS performance, in comparison with other state-of-the-art NAS methods, is summarized in the plot below. The X-axis shows how long (or expensive) the search process is (the lower the better), and Y-axis shows searched model’s performance.*